Competition for feature selection

Hannus, Aave

Chapter 1

Introduction

Perception is not something that happens to us, or in us. It is something we do.
Alva Noë, 2005, p. 1 (Noë, 2005)

While almost constantly confronted with an enormous amount of information, the human visual system appears to be highly efficient in differentiating between various aspects of this information on the basis of the situational demands. At any given time, only a small portion of the information available in the visual environment can be selected and identified for conscious processing. Optimally, this selection should be based on the information required for controlling current and planned behavior, and this process is described as *selective attention*. What are the determinants that decide which kind of information should be selected to enter conscious cognitive processing? The acknowledgement of the important role played by selective attention in the visual perception has established a long tradition of empirical and theoretical research starting from Helmholtz (1867) and James (1890) in the early days of experimental psychology. The aim of this thesis is to reveal specific interactions between some sensory and cognitive effects on attentional selection in visual cognition. To address those interactions, I will provide several basic principles and theoretical models of selective visual attention, and also the specific aims of the thesis. Firstly, I will describe attentional selectivity in visual cognition and its perceptual, cognitive, and neural origins. Secondly, the visual search as a behavior and also as a way to investigate visual cognition will be introduced. Thirdly, I will present a selection of relevant theoretical models describing sensory and cognitive mechanisms of selective visual attention. Finally, the framework and aims of the current studies will be provided.

1.1 Attentional Selection

Central to the entire discipline of cognitive psychology is the concept of attentional selection (Broadbent, 1958; Bundesen, 1990, 1998; Cherry, 1953; Deutsch & Deutsch, 1963; Duncan, 1980; Itti & Koch, 2001; Posner, Snyder, & Davidson, 1980; Treisman, 1960). Investigating the mechanisms that enable us to concentrate perceptual processing on some aspects of the environment and filter out other aspects of it is a continuing challenge within cognitive science (Clark, Squire, Merrikhi, & Noudoost, 2015; Desimone & Duncan, 1995; Driver, 2001; Moore, 2006; Moore & Zirnsak, 2017; Schneider, Einhäuser, & Horstmann,

---

Parts of the chapter are published as:

Hannus, A., Neggers, S. F. W., Cornelissen, F. W., & Bekkering, H. (2004). Selective attention for action: New evidence from visual search studies. In G. W. Humphreys & M. J. Riddoch, (Eds.), *Attention in action: Advances from cognitive neuroscience* (pp. 131–149). Hove: Psychology Press.

2013). Because of its structural and functional limitations, the human brain must process information selectively in a variety of domains. Specifically, we can restrict cognitive processing to a subset of the many potential visual or auditory objects that could be perceived and to a subset of the many potential actions that could be performed. Selective attention can be reflectively attraced by physical attributes or salience of objects, and it can be voluntarily directed toward objects of interest or some features of it. When someone wants to find, for example, her yellow bicycle in the large parking lot full of bikes, she can use her knowledge about the color, size, shape, the remembered location, or another particular feature of her bike to guide this search and locate her vehicle. If she decides to use the yellow color, then most probably she will serially track all yellow objects on the parking lot to figure out which yellow object contains the other characteristics of her bicycle. Indeed, experimental findings demonstrate that observers can restrict their visual search for a conjunction of color and orientation to a subset of the stimuli defined by color (Kaptein, Theeuwes, & van der Heijden, 1995).

The functions of visual perception are not limited to providing us with impressions of our visual surroundings. Much of the time, we need attentional selectivity not merely for selecting spatial locations, objects, or features, but to get things done—our visual system serves our behavioral intentions and action goals. For instance, I might want to find my bicycle in order to lift it out of the bike rack and mount it by putting my left foot on the left pedal, pushing off twice on the ground, swinging my right leg over, and sitting down on the seat in order to pedal away because I want to meet my friends on time. The existing body of research on selective visual attention suggests that relations between sensory input from visual environment and behavioral output from motor system are not unidirectional from perception to action but show a complex interaction between attentional processes and goal-directed behavior (for reviews, see Perry, Amarasooriya, & Fallah, 2016; Pratt, Taylor, & Gozi, 2015; Ridderinkhof, 2014; Schenk, 2010). The claim that attention acts to select the appropriate response according to the current behavioral goal received one of the most influential treatment in a seminal writing by Allport:

> Coherent, goal directed behavior requires processes of selective priorty, assignment, and co-ordination at many different levels (motivational, cognitive, motor, sensory). Together this set of selective and coordinative processes can be said to make up the effective attentional engagement (or attentional set) of an organism at any moment. (Allport, 1989, p. 652)

Given the knowledge about the complex selective visual attention, the obvious questions follow: how does the human brain select and interpret visual sensations to produce a visual percept, and what kind of methods allow us to assess,

evaluate and interpret visual perception and processing. In the following, I will describe the basic psychological functions and neural bases of selective visual attention and the research paradigm of visual search.

1.1.1 Control of selective visual attention: Stimulus-driven and user-driven processes

It is widely assumed that two types of processing guide selective visual attention. First, some objects attract our attention instantaneously, in a *bottom-up* manner. Such stimulus-driven control is the case when some internal attributes of the stimulus attract the attentional system, and the attention is captured reflexively because one particular stimulus is salient in a given context. The term *salience* is used to characterize the physical intensity of sensory stimulation in relation to surrounding stimuli (e.g., Itti & Koch, 2001). Thus, the bottom-up control of the visual attention represents sensory effects of physical properties of the visual stimulus and characterizes perceptual processes driven by sensory input. In contrast, in *top-down* or cognitively driven processing, the attention is guided voluntarily, based on the behavioral goals of the observer (Blaser, Sperling, & Lu, 1999; Carrasco, Ling, & Read, 2004). Therefore, top-down control of visual attention stands for the knowledge-driven processes, sometimes also called conceptually driven or user-driven processes, referring to the use of previous knowledge, experience, and intentions. Research on visual attention has been dominated by discussions about the roles played by bottom-up and top-down information in attentional processing (e.g., Anderson, Heinke, & Humphreys, 2012, 2013; Awh, Belopolsky, & Theeuwes, 2012; Beck & Kastner, 2009; Corbetta & Shulman, 2002; Hopfinger, Woldorff, Fletcher, & Mangun, 2001; Kastner & Ungerleider, 2000; McMains & Kastner, 2011; Theeuwes, 2010; van der Stigchel et al., 2009). This thesis aims to contribute to the knowledge about the interplay between these two sources of attentional modulation.

1.1.2 Neural basis of selective visual attention

The human visual system involves large subcortical and cortical brain structures. It has been suggested that in primates, approximately 55% of the cortex is specialized for visual processing (Felleman & Van Essen, 1991). However, as proposed by Desimone and Duncan in their influential review in 1995, the visual system is fundamentally characterized by its limited capacity for information processing.

Afferent visual system

The human visual system includes the eyes, connecting pathways through to the visual cortex, and large proportions of subcortical areas of the brain. Visual

perception starts in the retina, where the optical input is transduced. For the current thesis, it is important to stress that coding of color information starts already at the retinal level. Specifically, humans possess four types of photoreceptors: three types of cones and the rods. Under most visual conditions, the sensation is mediated by cones. Each type of cone has a unique, optimal response to particular wavelengths of light: short (blue), middle (green), or long (red). Differently, rods, which are saturated at natural light intensities and do not discriminate colors, are responsible for vision at low illumination conditions. The distribution of cones and rods across the retina reflects the different functions of the fovea and retinal periphery (Curcio, Sloan, Kalina, & Hendrickson, 1990; Østerberg, 1935). The *fovea* is located in the middle of the macula area of the retina to the temporal side of the optic nerve. As a result of its nearly 15-fold higher cone density compared to the peripheral retina, fovea provides high visual acuity (Hirsch & Curcio, 1989). Such variances in retinal cone density are the reason for making eye (and head) movements, which allow to bring the foveal region of the eye on top of the area of interest and maximize visual processing resources in that particular area of the visual field.

Photoreceptors deliver visual information to bipolar cells, which relay it to ganglion cells. Neural signals further travel via ganglion cells through the optic nerves, dividing and partially crossing over into the optic chiasm and then going via the optic tracts to the dorsal part of the lateral geniculate nucleus (LGN) in the thalamus. The LGN is the first central relay through which visual information passes on the way to the cortex. From the LGN, the neural signals continue to the primary visual cortex in the occipital lobe, where further visual processing takes place. Also, the optic nerve sends a branch to the superior colliculus (SC) that regulates orientation movements of the eyes and the head. However, SC receives inputs from several brain areas, most notably from the visual cortex and frontal eye fields (FEF) in the premotor cortex. Importantly, SC also connects to midbrain and brainstem, where the retinotopic representations of visual objects are transformed to motor programs.

In the visual cortex, a total of six distinctive areas is known: V1, V2, V3, V3a, V4, and V5. The neurons from the LGN synapse in the primary visual cortex V1 or striate cortex, where the neural signals are interpreted in terms of orientation, luminance contrast, spatial frequency, direction of motion, and color (e.g., Carandini et al., 2005; De Valois & De Valois, 1993; Horwitz & Hass, 2012; Hubel & Wiesel, 1959). As the neural signals continue further into higher areas of the visual cortex, more associative processes take place. The primary visual cortex projects to other regions of the cerebral cortex that are involved in complex visual perception. Adjacent extrastriate visual areas, each specialized for the detection of particular visual attributes, are organized into two roughly parallel processing streams. Specifically, in 1982 Ungerleider and Mishkin proposed the distinction of processing of different kinds of visual information in

the inferior temporal and superior parietal cortex accounting for appreciation of object's qualities ("what") and of its spatial location ("where"), respectively. Thus, the *dorsal visual pathway* was suggested to mediate the location of visual objects, while recognition and identification of those objects was attributed to the *ventral visual pathway*.

The dorsal pathway, also called occipitoparietal stream, begins in the area V1, projects to the thick stripes in areas V2 and V3, and passes to the area V5 and to the posterior part of the inferior parietal lobe (Mishkin, Ungerleider, & Macko, 1983; Ungerleider & Mishkin, 1982). More recently, Kravitz, Saleem, Baker, and Mishkin (2011) have also described the anatomical and functional trifurcation of the dorsal stream beyond the parietal cortex into prefrontal, premotor, and medial temporal areas. The ventral pathway or occipitotemporal stream also begins in the area V1 and projects to the thin and interstripe regions of V2, mainly representing color and object form, and projects to the area V4, which ultimately connects to the inferior temporal cortex (Mishkin et al., 1983; Ungerleider & Mishkin, 1982; for a more recent review, see Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013). The fundamental division of labour between the two visual streams has suggested that dorsal pathway provides mainly spatial information and object features for the planning and programming of motor actions (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). Differently, the ventral pathway has been suggested to be responsible for conscious visual perception, recognition, and construction of long-term representations from object features and their relations.

Initially, differentiation between the visual streams was primarily based on evidence derived from animal studies which indicated that inferotemporal lesions impair the identification of visual objects, while lesions in the posterior parietal areas cause visuospatial deficits. However, in 1991 the preeminent case of the patient D.F. suffering a damage of lateral occipital and parasagittal occipitoparietal regions was introduced (Goodale, Milner, Jakobson, & Carey, 1991). Such a damage usually leads to visual form agnosia that is manifested by a severe deficit in the perception of size, shape, and orientation of visual objects. This deficit was also described for D.F. Importantly, when D.F. was asked to make reaching movements and orient her hand or to pick up a block placed at different orientations in front of her, her aiming and prehension performance was correct. Thus, evidently, D.F. could use information about object orientation accurately for visuomotor action, but she was unable to use the same information for perceptual purposes. Later neuroimaging has confirmed D.F.'s damage to the ventral stream and associated her spared visuomotor functions with visual processing in the dorsal stream (James, Culham, Humphrey, Milner, & Goodale, 2003). Conversely, optic ataxia resulting from unilateral lesions of mostly posterior parietal areas induces visuomotor deficits in reaching and grasping for objects, while recognition and matching of objects are not affected.

Specifically, damage to the parietal cortex can lead to deficits in the positioning of fingers and in adjusting of the hand during reaching movements (Jackson et al., 2009; Milner & Goodale, 1995; Perenin & Vighetto, 1988).

Having established the functional distinction between ventral and dorsal visual streams or "vision for perception" and "vision for action " does not imply isolated segregation of the two pathways. Indeed, recent evidence indicates that the strict distinction between two streams of visual processing or their supposed differential functions is not successful (de Haan & Cowey, 2011; Freud, Plaut, & Behrmann, 2016; Goodale, 2014). Milner and Goodale (2008) have highlighted that the division of labour between ventral and dorsal visual streams serves different metrics and frames of reference for perception and action. In the more recent view, the ventral visual stream has evolved for scene-based representation of the properties of an object in relation to other objects and this information is also used for *movement planning*, whereas the dorsal stream implements egocentric computation of object's spatial properties for *programming* and online *control of action* (Goodale & Wolf, 2009; Milner & Goodale, 2008). The two visual streams reveal connections at several levels. For instance, intermingled projection cells have been found in V4 and at the bottom of the anterior superior temporal sulcus (Baizer, Ungerleider, & Desimone, 1991). In a delayed reaching task, optic ataxia patients have indicated a delay-related gradual change between dorsal and ventral control of reaching behaviour rather than a discrete switching between the two pathways (Himmelbach & Karnath, 2005). Similarly, some authors have challenged the simple double-dissociation between visual form agnosia and optic ataxia and describe more complex multiple parallel substreams of visuo-motor control (Pisella, Binkofski, Lasek, Toni, & Rossetti, 2006; Pisella et al., 2009). Recently, it has been argued that skilled and accurate grasping movements rely on interactions between dorsal and ventral pathways (van Polanen & Davare, 2015). Coordinated recruitment of ventral and dorsal modulatory signals is probably achieved by combined bidirectional cortical projections. In a neurophysiological animal study, a causal contribution of a dorsal stream area to cortical activation in the ventral stream has been demonstrated (Van Dromme, Premereur, Verhoef, Vanduffel, & Janssen, 2016). Moreover, a human neuroimaging study has shown that the vertical occipital fasciculus connecting occipital, temporal, and parietal cortex is likely the crucial link between dorsal and ventral streams (Takemura et al., 2016).

Attentional control system

Attentional control of sensory processing and selection is implemented in a large network comprising frontal and parietal areas that integrate bottom-up and top-down priorities. Both spatial and object-based shifts of attention are accompanied by transient increases in activation of regions in superior pari-

etal lobule (Yantis et al., 2002; Yantis & Serences, 2003). In addition, as suggested by Miller and Cohen (2001), the higher-order cognitive control of selective attention appears to be largely achieved by multimodal convergence and integration of neural activity in prefrontal cortical areas. Although attention modulates neuronal response in extrastriate cortical areas encoding specific visual features (Maunsell & Treue, 2006), for instance, attending to color enhances response of color-sensitive cortical regions (Chawla, Rees, & Friston, 1999; Liu, Slotnick, Serences, & Yantis, 2003; Saenz, Buracas, & Boynton, 2002) and action preparation modulates neural activity related to action relevant visual features as early as in V1 (Gutteling et al., 2015), it is now well recognized that visual search tasks activate large proportions of the dorsal frontoparietal cortex (for review, see Corbetta & Shulman, 2002; Kastner & Ungerleider, 2000; Ptak, 2012; Scolari, Seidl-Rathkopf, & Kastner, 2015). Specifically, *frontoparietal attention network* (FPAN) refers to dorsal regions that are concurrently activated during attentional shifts and encode the priorities in the representation of the environment. The corresponding dynamic topographic organization of spatial representations of visual features and observer's intentional biases is called priority mapping. Particularly, it has been suggested that stimulus-driven salience maps configured by the conspicuousness of bottom-up information from early visual neurons about the physical stimulus properties (cf. Itti & Koch, 2001) are combined with behavioral salience maps relying on action-related top-down information from higher association areas in order to form priority maps that guide visuomotor behavior (Fecteau & Munoz, 2006; Serences & Yantis, 2007). The FPAN comprises strongly interconnected areas in the posterior parietal, premotor, and prefrontal cortex, including also the FEF.

A detailed overview of the massive feedforward and feedback projections of the neural basis of priority mapping is beyond the scope of this chapter. The most evident indices show that stimulus-driven salience is initiated from early visual areas and subcortical regions, while learned value-based maps are triggered from limbic structures in the midbrain, and top-down instruction-based salience originates from frontal cortex; the combination of these three sources of attentional modulation is feed to the FEF that control eye movements (Klink, Jentgens, & Lorteije, 2014). Evidence from functional connectivity examination suggests functional interactions between ventral and frontoparietal cortical regions and indicates that lateral occipital areas mediate coupling between the ventral and dorsal pathways during sensorimotor tasks (Hutchison & Gallivan, 2016).

Ocular motor system

The human ocular motor system has two primary functions (for review, see Kowler, 2011; Spering & Carrasco, 2015). Firstly, some eye movements have to

keep the eye fixed on objects of interest. When the image of a stationary object needs to be held, *fixations* keep the eye fixed on the object. Also, *vestibular eye movements* help to maintain clear vision when the head is moving, and *optokinetic eye movements* contribute to maintaining clear vision when the visual stimuli are moving with respect to the head. Secondly, since visual acuity is greatest when an object is positioned on the fovea, *saccadic eye movements* allow moving the eyes toward objects of interest rapidly, and they are characterized by the consistent relationship between peak velocities and amplitudes (Smit, van Gisbergen, & Cools, 1987). In addition, thirdly, smooth *pursuit eye movements* help to hold the image of a moving object on the fovea. Finally, due to the frontally positioned eyes, humans need to use *verging eye movements* to keep the foveae of both eyes on objects of interest.

1.2 Visual Search: Exploratory Behavior and Research Paradigm

A central research paradigm that has been developed to study the characteristics of selective visual processes is known as *visual search*. This paradigm exploits a very basic exploratory search behavior that has attracted a lot of interest in both applied and fundamental research for more than 100 years (Nakayama & Martini, 2011). Experimental visual search tasks require observers to search for a pre-specified *target* among an array of nontarget distractors, thereby allowing to explore deployment of visual attention (Posner, 1980, 1992; Treisman & Gormican, 1988; Wolfe, 1994). A typical visual search task starts with visual input and ends with output, usually in the form of some behavioral response indicating that the target (or its absence) was detected. Visual search experiments aim to reveal mechanisms of visual attention by time locking visual events to either overt shifts in eye position (i.e., saccades) or covert orienting responses. Therefore, two different types of visual search tasks exist (for a review, see Findlay, 2003). The more frequently used task is the *covert search* task that requires observers to use covert attention, i.e., direct their attention to some part of the stimulus array without moving their eyes. Research on covert attention has yielded the traditional metaphors of visual attention functioning as either a spotlight that *"enhances the efficiency of the detection of event within its beam"* (Posner et al., 1980, p. 172) or a *zoom lens* that distributes processing resources evenly over the zoomable area of attentional focus (Eriksen & St James, 1986). Less frequently, the *overt search* tasks are used where observers are invited to apply overt attention and move their eyes in order to align the fovea with an object in the stimulus array. Thus, overt search tasks require the observer not only to decide about the presence or absence of a particular object but are further interested in the pattern of eye movements that the observer undertakes in the way of the search. It is assumed that observers fixate on one point of the display and use

the peripheral vision to decide which spatial location would be the most relevant for the next fixation (Bloomfield, 1979; Williams, 1966). This method assumes that the decisions to sequentially foveate further areas of the display reveal the underlying attentional processing. Thus, the measurement of the paths of eye movements and fixations can give information about how the visual attention is deployed.

As mentioned above, attention is selective. Visual search studies have aimed to expose fundamental principles of attentional selectivity. One key issue about attentional selectivity is the stage of information processing: at what level of processing does this selectivity start to operate? According to *early selection* models, the role of attention is to filter out some information at a very early stage and enable only attended information to reach cognitive processing, whereas unattended information undergoes only rudimentary processing. Early selection is thought to operate through two systems. In preattentive parallel processing, simple physical characteristics of stimuli are extracted (e.g., their color). The attentive system, characterized by limited processing capacity (i.e., the entire visual scene cannot be processed at once, but rather piece by piece), processes only stimuli passed from the first, preattentive analysis; the categorical identity of stimuli should be processed at this later stage of processing. A typical example of the early selection approach is Broadbent's (1958) filter theory. In contrast, *late selection* models propose that unattended stimuli are not rejected from full processing, but only from entry into memory processes or the control of voluntary responses (Duncan, 1980). It is assumed that already on the first, parallel level of processing, a full analysis and extraction of semantic categorization of the stimuli is performed, and this information is used as the basis for selection for the system with limited processing capacity. However, it is important to stress that often in the literature there appears to be no systematic differentiation between selective cueing and selective processing. Allport has claimed that typically, early selection echoes the selective cueing or specifying of task-relevant information and late selection characterizes further processing of both relevant and irrelevant information (Allport, 1987, 1989).

The majority of theories describing and explaining visual search and selection deal with the interplay between bottom-up and top-down information. To explain it in detail, we also need to elaborate on parallel and serial processing. The cognitive processing is termed *parallel* if all elements are processed at the same time, assuming that processing of all entities starts simultaneously over the entire visual field. Alternatively, with *serial processing*, the elements are processed one by one, with the processing of one element being completed before starting the processing of another (Townsend, 1971). Initially, Treisman and Gelade (1980) suggested that visual search for targets defined by unique salient features could be performed in parallel: the target is effortlessly located on the basis of preattentive visual processing over the visual scene, which means, that

it "pops out" from the visual array and can be found almost immediately (the response time does not depend on the number of objects; see Figure 1.1). Conversely, searching for a particular combination or *conjunction* of features leads to a less efficient serial search. However, as later suggested, the parallel and serial search patterns appear to represent the two extremes of a continuum of search results rather than indicate the existence of a strict dichotomy (e.g., Wolfe, 1996). The search efficiency can be explained by stimulus similarity rather than by distinctive parallel and serial processing (Duncan & Humphreys, 1989). Pashler (1987) has suggested a molar serial self-terminating search over clumps of stimuli, while within these clumps a capacity-limited parallel search takes place; the size of clumps appears to be around eight items. Recently, Buetti and colleagues have provided evidence for the principle of unlimited capacity parallel processing whereby they showed that even parallel visual search efficiency is logarithmically dependent on the set size and also sensitive to stimulus similarity (Buetti, Cronin, Madison, Wang, & Lleras, 2016).

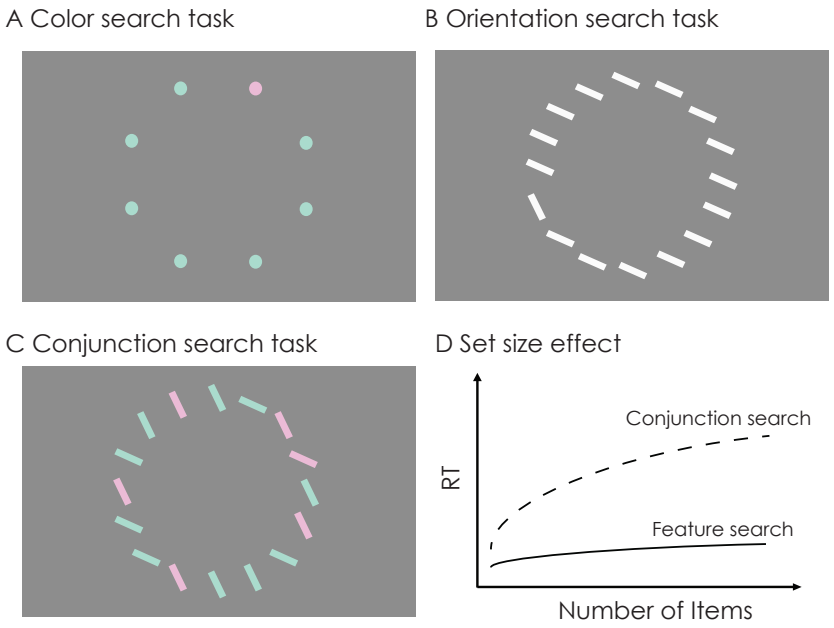

Figure 1.1: Visual search paradigm. In a typical visual search task, observers are required to determine if a predefined target is present among nontarget distractors. Usually, the number of items (set size) is systematically manipulated. The conventional outcome measures are the target detection or discrimination accuracy and reaction time (RT) that is needed to make a response decision. A. Sample display from a singleton color search task, set size 8 items. B. Sample display from singleton orientation search task, set size 16 items. C. Sample display from a conjunction search task of color and orientation, set size 16 items. D. Conceptual illustration of the set size effect indicating that singleton search is relatively unaffected by set size whereas in conjunction search the RT depends on the number of search items.

While behavioral detection and discrimination tasks are frequently used to assess covert visual search parameters, a standard part of the psychophysical research on overt visual search is eye movement recording (Figure 1.2). The link between eye movements and visual perception is so tight that visual processing is enhanced at the very early levels of eye movement preparation (Kowler, Anderson, Dosher, & Blaser, 1995; Rolfs & Carrasco, 2012). Therefore, in the present thesis, I used the saccadic decision making to address the visual selection processes. All experiments presented in this thesis applied only two dependent measures: (a) landing point (location) of the initial saccade after the onset of a particular visual search display and (b) saccadic latency of the initial saccade. Next, I will briefly describe the basic principles of some theoretical approaches that appeared most relevant for the current thesis.

1.3 Modulation of Selective Visual Attention

1.3.1 Feature integration theory

A very influential early model of selective visual attention was the feature integration theory (FIT; Treisman, 1977, 1991; Treisman & Gelade, 1980; Treisman

Figure 1.2: Video-based eye tracker ASL 5000 Series, Model 501 (Applied Science Laboratories, Bedford, MA, USA) used in the experiments presented in Chapters 2 and 3. Note that this experimental apparatus, as combined with the electromagnetic position tracking system (miniBIRD 800TM, Ascension Technology Corporation, Shelburne, VT, USA), allowed participants to make free head movements and therefore quite naturalistic hand movements toward visual objects (Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, NL).

& Sato, 1990) that proposed the existence of two qualitatively different search mechanisms: *single search* for a target defined by basic visual feature dimensions and *conjunction search* for a target defined by a combination of features (Figure 1.1). Specifically, searching for *singletons* (i.e., single features such as color, orientation, or shape) is much less effortful than searching for targets defined by conjunctions of two stimulus dimensions (Treisman & Gelade, 1980). The search effort has been measured as the dependence of reaction time (RT) on the number of nontargets, or the number of errors per *set-size* (number of stimuli in the visual array). According to FIT, singletons can be detected by a preattentive parallel search without fully identifying distractors. FIT has suggested that singletons are identified as an activity in independent *feature maps* tuned to discrete features and allowing for spatially parallel search processes. Importantly, it appeared that in such single search tasks the function of RT in relation to the number of nontargets is flat or almost flat, i.e., the set size effects are around 6 ms per object or less (Treisman & Souther, 1985). The fast target detection in the single search tasks was called *pop-out*, as the salient singleton tends to capture attention in a bottom-up manner. Alternatively, the detection of targets defined by feature conjunctions usually involves the much slower process of serial scanning of all the elements in the array. FIT suggested that detection of a conjunction requires focal, top-down modulated attention, which is serially directed and permits accurate localization and binding of several features. Thus, according to FIT, the identity of elements in a visual scene ("what") and their location ("where") are unified by a serial scan of spatial locations using a window of attention. In this manner, the features of the visual scene are bound and then compared to stored representations of objects for recognition (Treisman, 1996; Treisman & Gelade, 1980; Treisman & Gormican, 1988). Therefore, attention is thought to serve as a binding mechanism for gluing up the presumably separately represented features belonging to a particular object and for making an overall decision about the representation of this stimulus—e.g., is it likely to be the target stimulus to foveate for further inspection.

FIT has received extensive experimental support. There is evidence for the parallel processing in a single search. The features leading to pop-out effect are color, size, shape, orientation, and curvature (e.g., Treisman & Gelade, 1980). It has been shown that similarity among nontargets determines the strength of perceptual grouping, while similarity between target and nontargets determines the ease of segregating target from nontargets (Duncan & Humphreys, 1989; Humphreys & Muller, 1993). However, it is important to note that the revised version of this model suggests that spatial attention might be applied for the detection of stimuli defined by both features and conjunctions (Treisman & Sato, 1990). A remarkable amount of findings suggests that the strict dichotomy between parallel and serial search is not justified and does not necessarily reflect the essential differences between feature and conjunction search (e.g.,

Duncan & Humphreys, 1989; Eckstein, 1998; Findlay, 1997; Pashler, 1987).

1.3.2 Guided search model

Although influenced by FIT, the guided search model (GS; Wolfe, Cave, & Franzel, 1989) has proposed that single features and a feature conjunctions are not processed differently. According to GS, massively parallel cognitive processes guide attention to likely targets but do not distinguish whether an item is a target or not. GS posits that basic visual features are detected across the retina in parallel, yielding a set of feature activity maps that represent each feature dimension via a coarse coding. The maps for the feature dimensions are overlapping and broadly tuned. For instance, color might be represented by maps tuned to red and green, and orientation might be represented by maps tuned to steep and shallow-sloped edges. The feature activity maps are suggested to pass through a differencing mechanism that delivers a bottom-up activation. Therefore, parallel processes are thought just to divide the set of stimuli into candidate targets and those that could most likely not be a target. Subsequently, a serial process is used to search through the selected portion of the visual field until the target is detected. In this way, the "spotlight" of attention is guided by information from parallel processes. All feature maps deliver bottom-up activations that are combined into a *salience map*, where local activities indicate the priority of a particular location for the current task. Notably, triple conjunction search tasks appear even easier than detection of conjunctions of two features (the effect of set size on RT is weaker) because the ongoing parallel processes provide relatively more information to the serial process (Wolfe et al., 1989). According to GS, if the signal from a stimulus containing feature value of the target is strong enough, with a higher probability the attention is guided to this particular stimulus. In this way the top-down information can be effectively used for guiding the focal attention to those stimuli that contain relevant feature characteristics delivered by bottom-up processes; physiological correlates have supported this model (e.g., Luck & Hillyard, 1994).

Wolfe has further refined his model to make it more realistic accounting for visual search processes in real word tasks and to integrate the eye movements into the model (Wolfe, 1994, 2007; Wolfe & Gancarz, 1996). In those recent developments—GS2, GS3, and GS4—attention acts in a manner that allows only features of one object to reach the higher visual processing at a time. Consequently, the activity map—the weighted sum of activity in the preattentive feature maps, working by a winner-takes-all principle—is assumed to guide visual attention.

1.3.3 Similarity theory

In contrast to FIT, a strict dichotomy between parallel and the serial search was

challenged by Quinlan and Humphreys (1987). Specifically, they showed that search time might depend on the amount of information required to identify the target stimulus, and also on the target-nontarget discriminability. Accordingly, Duncan and Humphreys (1989) suggested that at the first, unlimited parallel stage of the processing the visual representation of stimuli is segmented into *structural units*, for instance, based on proximity or similarity. These units were thought to form a perceptual description of the visual input representing the structure of this information. The similarity theory proposed that individual structural units are organized hierarchically, consisting of sets of properties like color, shape, or motion, whereby each structural unit may further be segmented into smaller units. In this way, a hierarchical representation of the visual field should be produced. Next, it was proposed, the input descriptions are compared to the internal template of the target, whereas the structural units containing some property of template can get a higher weight and thus a higher chance to get selected (or to enter the capacity-limited visual short-term memory). At this moment, the attention could be directed to some aspects of incoming information, e.g., orientation or color of structural units. This theory suggests that visual search efficiency depends on dissimilarity between targets and nontargets, and similarity between nontargets (Duncan & Humphreys, 1989). Specifically, since structural units are hierarchically grouped, a poor match between the visual template and a structural unit allows efficient rejection of other units that are strongly grouped. To sum up, similarity theory allows to make predictions about the course of the visual search, but unfortunately not about the interactions between different features.

1.3.4 Biased competition model

In 1995 Desimone and Duncan proposed a further model of visual selection, the biased competition model (BCM; Desimone & Duncan, 1995; Reynolds, Chelazzi, & Desimone, 1999). Their account was based on the interplay between both bottom-up and top-down sources of attention. According to BCM, features of an attended object are processed concurrently, but the limitation of the ability to deal simultaneously with several sources of visual information determines the number of separate objects that can be processed. Due to these constraints in attentional capacity, a selective system should operate to restrict the huge amount of potential inputs and withhold information irrelevant to the current behavior. BCM suggests that among visual objects a competition for representation and analysis takes place. Specifically, it is assumed that mutually suppressive interactions between competing stimuli facilitate attentional selection and preactivated target units have an advantage in this competition (Duncan, 1996). It has been proposed that within the brain systems receiving visual input, a gain in activation for one object entails a loss of activation for other objects.

For instance, Duncan (1984) indicated that two attributes of a single object (e.g., color and orientation) could be identified simultaneously without mutual interference, while attributes of two different objects could not, even if the objects spatially overlap. This object-based theory suggests that focal attention is guided by parallel, preattentive processes representing discrete objects. One plausible cause for this competition might be the structure of cortical areas in both the ventral and dorsal visual stream (for a more detailed discussion, see Beck & Kastner, 2009; Kastner & Ungerleider, 2001). Biased competition is attributed to the fact that as the complexity of visual processing increases in every consecutive cortical area, receptive field sizes of individual neurons increase. When more objects are added to the receptive field, the information about any given individual object must decline. Desimone and Duncan (1995) proposed competition between objects represented by the same receptive field. In the context of the present thesis, it is important to stress that BCM recognizes the role of action-relevant top-down influences as a reason for biased suppression: the competition is biased toward information relevant for the current behavior. The information in a visual scene determines the spatial distribution and feature attributes of objects. During the search in this information, a target may pop out due to a bottom-up bias that directs attention toward local inhomogeneities. On the other hand, selectivity is implemented to bias the competition toward behaviorally relevant information using top-down control (Beck & Kastner, 2005, 2009).

Importantly, the further development of theories of selective attention reflects a gradual shift from merely perceptual considerations toward a stronger emphasis on the interaction between sensory and perceptual processing and behavioral actions. Next, I will introduce the concept of selection-for-action along with a few far-reaching studies that have clearly differentiated perceptual and action-related effects in selective visual attention.

1.3.5 Selection-for-action

In his 1989 writing, Allport pointed out that since the 1950s the majority of research in the area of visual attention had mainly considered the limited information-processing capacity of the brain as the fundamental constraint underlying all operations of attention. Thus, according to the earlier views, the selectivity function of attention arose from the limited capacity of information processing system. Correspondingly, experimental psychology had been dominated by attempts to explain the functions of visual attention as "selection-for-visual-perception" (Deubel, Schneider, & Paprotta, 1998) that resulted in the theories and models described above. Diverging from the general idea that attention operates as a mechanism for coping with central limited capacities of cognitive processing, Allport stressed the constraints in preparation and control of

action (Allport, 1987, 1989). The idea behind proposing the selection-for-action perspective was that integrated actions require the selection of particular aspects or attributes of the environment that are relevant to this action at hand, whereas the information irrelevant to the action should be ignored. Thus, the attentional processes are viewed as the selection of action-relevant events or stimuli relying on particular action plans. The basis of this approach can be linked to the work of Lotze, whos idea of the ideomotor principle (Lotze, 1852) proposed that actions are selected and planned regarding their sensory consequences. In other words, perception can be directly linked to upcoming action intentions since both are represented in the brain in sensory terms. Therefore, the attentional processing might reflect the necessity of selecting information relevant to the task at hand.

Reasons for this kind of argumentation had also emerged from the notion that selecting a stimulus as a target for a saccade occurs before the foveation. The efficiency of covert attention has suggested that foveation is neither necessary nor a sufficient condition for selection (Allport, 1987). However, both top-down saccadic selection for visual perception and selection for motor actions are coupled to the target object. Specifically, in their seminal study published in 1996, Deubel and Schneider demonstrated the inevitable coupling of ventral processing for perception and dorsal processing for saccade programming being restricted to one common target object at a time. By showing increased visual target discrimination due to pointing to the same target, Deubel and colleagues (1998) further confirmed analogous coupling of attentional selection and goal-directed hand movements. As a result, a strict one-object-at-a-time rule was proposed suggesting that goal-directed action toward an object relies on perceptual processing of the movement target (Deubel et al., 1998). Taken together, object perception requires binding of information about different attributes of that object, which then allows the purposeful use of the object according to intended action. For instance, if the intention is to take a yellow dictionary out of the bookshelf, the information eventually about its color, size, and orientation should be combined to execute an accurate grasping movement. Importantly, however, a specific action intention, such as grasping an object, can selectively enhance visual processing of action-relevant features, such as the orientation of the object (Bekkering & Neggers, 2002).

1.3.6 Theory of event coding

Interactions between action-relevant and stimulus-driven effects on selective attention are comprehensively described by Hommel and colleagues in the theory of event coding (TEC; Hommel, 2009; Hommel & Colzato, 2009; Hommel, Musseler, Aschersleben, & Prinz, 2001). Similarly to the selection-for-action model, TEC emphasizes perceptual consequences of action in goal-directed

human behavior. TEC is founded on the idea that a voluntary action is cognitively represented by the code of its *perceptual consequences* and the bidirectional associations between motor patterns and movement-contingent events are formed by learning (Elsner & Hommel, 2001). As a matter of fact, according to TEC, representations of perceived events and produced actions are considered essentially the same and are cognitively represented in common *event codes* (Hommel, 2009; Hommel et al., 2001). Further, TEC assumes that event codes are composed representations of feature codes or cognitive correlates of visual and motor features. Therefore, it is further assumed that perception and action share the basic units called sensorimotor entities that are activated both by sensory input and motor control. TEC suggests that activation of an action plan—whether due to action goals or stimulus properties—increases weights of specific perceptual features subserving perception, action planning, and action adjustments. Therefore, according to the TEC, it is possible to increase the weights of a particular feature to facilitate the coding of this feature. For instance, the action-related anticipation of the behaviorally relevant feature would increase the weights of that feature and therefore enhance its processing. Accordingly, as suggested by TEC, the intention to grasp an object primes object's orientation in space and therefore facilitates orientation processing. In this manner, the mere activation of an action plan could stimulate certain intentional weighting mechanisms and thereby increase the weights of those features that allow for the specification of action parameters (Hommel, 2010).

The studies presented in this thesis are devoted to furher demonstrate the complex interplay between stimulus-driven propeties of visual objects and action-relevant goals of human actors.

1.4 Issues and Outline of the Thesis

The four studies presented in this thesis are dedicated to exploring interactions between top-down and bottom-up influences on selective visual attention. This thesis was inspired by a remarkable study on the effects of action intention on selective visual attention (see Figure 1.3). Specifically, Bekkering and Neggers (2002) found that when observers had to grasp a target stimulus combining color and orientation, a significant improvement in orientation discrimination accuracy appeared in comparison to the condition where they had to point to the target. Notably, the discrimination accuracy of color was not related to the type of hand movement. Since the relative orientation in space was assumed to be more important for the grasping preparation than for the pointing preparation, the authors suggested that the action-relevant visual feature can selectively be processed more efficiently than the action-neutral feature. The present thesis aims to further explore the top-down induced effects on selective visual attention and test if the action intention has an effect only on the action-relevant

feature dimension or does it bias the competition between action-relevant and neutral features.

The traditional models of visual selection described above predict that in effortful conjunction search, the top-down activation will determine the course of the search. For example, Yantis and Jonides (1990) presented an elegant demonstration of the power of voluntary processes over reflexive attentional capture. They showed that if observers know with certainty the location of a target and have sufficient time to focus attention on that target, the nontarget distractors are at least temporarily unable to capture the attention. Other researchers have claimed that the attentional capture does not occur due to the stimulus or nontarget properties per se, but is determined by the relationship of nontarget properties to target-finding properties. Specifically, it has been proposed that the involuntary attentional capture is subject to top-down control and occurs if and only if nontargets have a property that the observer is using to find the target (Folk, Remington, & Johnston, 1992). On the other hand, research conducted by Theeuwes and his colleagues has pointed out the limitations of top-down modulation visual attention (for review, see Theeuwes, 2010, 2013; van der Stigchel et al., 2009). However, further research on naturalistic stimuli suggests that visual search can override bottom-up salience, and top-down signals modulate bottom-up effects in a feature-specific manner

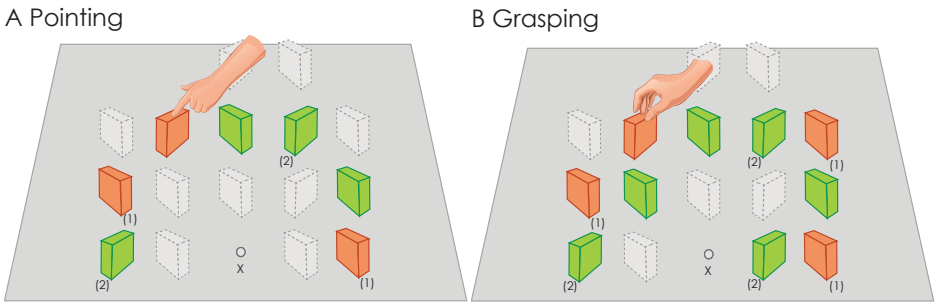

Figure 1.3: Schematic of the experimental paradigm used by Bekkering and Neggers (2002). Experiment took place in a totally dark room. At 16 possible locations, randomly, either one, four, seven, or ten objects were illumnated by the LEDs below them, one of the objects was the predefined target. Participants were instructed to gaze at the fixation dot (X) until they thought that they know the location of the target. After that, in the pointing condition (A), they were required to point to the centre of the top surface of the target, in the grasping condition (B), to grasp the target between the thumb and index finger. One third of the nontargets had the same color as the target, 1/3 had the same orientation as the target, and 1/3 had both a different color and orientation. In the present examples, the target is the orange clockwise oriented object, nontargets are orange objects oriented counterclockwise, green objects oriented counterclockwise, and green objects oriented clockwise; an initial saccade to (1) would be considered an orientation error and an initial saccade to (2) would be a color error. Achromatic objects represent possible object locations that are not illuminated during these sample trials. Adapted from "Visual search is modulated by action intentions" by H. Bekkering and S. F. W. Neggers, 2002, *Psychological Science, 13,* pp. 371-372.

(Einhauser, Rutishauser, & Koch, 2008). Clearly, much uncertainty still exists about the interactions between action-relevant and stimulus-driven sources of attentional modulation. In the following chapters, I will present four studies indicating competitive interactions within the stimulus-driven bottom-up information, and also between action-related and stimulus-driven effects directing visual search behavior. The studies are ordered thematically, reflecting the two lines of experiments that allow drawing conclusions about the action-related and stimulus-driven effects on visual selection.

Chapter 2 introduces two experiments manipulating action-related manual tasks and stimulus-driven visual properties. The results of the first experiment showed that when one feature is selectively relevant for the planned action, the discrimination accuracy of this feature increases. In this study, we demonstrate a competition between the color and orientation and show that this competition can be biased toward orientation if the orientation is relevant for the manual action at hand. Importantly, we also showed that this effect is sensitive to the set size and disappearing when more cognitive effort is required. In the second experiment, the color discriminability was lowered, increasing the effort to be invested into color processing. By this manipulation, the top-down action intention effect disappeared. This finding leads to the conclusion that the action intention does not selectively enhance processing of the behaviorally relevant feature, it rather biases the competition between the conjunction features.

Chapter 3 further explores the top-down modulation of visual selection due to action intention. Specifically, we found additional evidence for competition between different feature dimensions. The results showed that behavioral intention related to the manual task could also decrease color discrimination accuracy. Particularly, in a color singleton search task where all stimuli had the same orientation, color discrimination accuracy decreased when orientation was relevant to the manual action at hand. Thus, although the stimulus-driven information represented a typical color singleton search task, the top-down intention to grasp the colored but also oriented stimulus biased the competition between the features away from color.

Chapter 4 explicitly addresses the question whether the visual feature dimensions are processed independently from each other or whether do they rather interact. Three experiments demonstrated that in conjunction search the features are not processed independently from each other. Specifically, we equalized the feature discriminability of color and orientation or color and size in singleton search tasks. After that, the a priori equally salient color and orientation or color and size were combined to form conjunctions. The findings showed that in conjunction search observers tend to initiate their search by saccading to a stimulus with target color, while the accuracy of orientation and size discrimination substantially decreases. Thus, if there is an equal amount of stimulus-driven information about color and orientation or color and size, color

becomes the prioritized feature.

Chapter 5 further confirms a biased competition taking place between conjunction features. In two experiments we provided prior information about either the color or the orientation of the subsequent conjunction and estimated the resultant feature discrimination accuracy relative to the not precued conjunction search performance. It appeared that if information about the orientation values of the subsequent conjunction stimuli is precued for a relatively long time, observers can use this prior information and increase subsequent orientation discrimination accuracy, while at the same time color discrimination accuracy decreases. Differently, when a color precue with a particular (but non-informative) orientation had been presented, the subsequent relative gain in orientation discrimination accuracy was equal to the relative gain in color discrimination accuracy. While the strong stimulus-driven bias toward color discrimination that appeared in Chapters 2 and 4 was confirmed, we also showed that a sustained top-down bias induced by orientation precueing could reverse the bottom-up bias toward color discrimination.

On the basis of our findings presented in Chapter 4 and 5, I suggest that in conjunction search of color and orientation (and also color and size) the bottom-up activation is higher for color than for orientation. Why this bias arises remains somewhat of a question, but could be related to the use of conjunctively tuned visual channels during conjuction search. Importantly, this finding cannot be directly predicted from the traditional visual search theories referred to in Section 1.3. Specifically, theories and models presented above do not assume one feature to be more effective in conjunction search when the stimulus-driven saliences are matched at the feature level. Therefore I suggest that the traditional views need an additional specification about this asymmetry in the feature processing. Findings presented in the following chapters suggest a competition between the features in conjunction search, and this is a priori won by color. The original BCM describes suppression of representations of the task-irrelevant visual objects and has excluded competition between individual features (Duncan, 1996; Duncan, Humphreys, & Ward, 1997). However, contrary to this assumption, other observations indicate that the suppressive competition could take place not only among nearby objects but at the level of individual features within an object (Beuth & Hamker, 2015; Haenny & Schiller, 1988; Martinez-Trujillo & Treue, 2004; Motter, 1994; Polk, Drake, Jonides, Smith, & Smith, 2008). The present thesis suggests that biased competition does not only operate at the level of objects—increasing the relative weight of one object at the cost of others—but it also serves early object segregation by biasing the relative weights of the features of visual objects.

Chapter 6 provides a summary of all findings and discusses the results in a broader framework of their theoretical implications.